

VU Research Portal

Physical and Cognitive Functioning After 3 Years Can Be Predicted Using Information From the Diagnostic Process in Recently Diagnosed Multiple Sclerosis

de Groot, V.; Beckerman, H.; Uitdehaag, B.M.J.; Hintzen, R.Q.; Minneboo, A.; Heymans, M.W.; Lankhorst, G.J.; Polman, C.H.; Bouter, L.M.

published in

Archives of Physical Medicine and Rehabilitation
2009

DOI (link to publisher)

[10.1016/j.apmr.2009.03.018](https://doi.org/10.1016/j.apmr.2009.03.018)

document version

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

citation for published version (APA)

de Groot, V., Beckerman, H., Uitdehaag, B. M. J., Hintzen, R. Q., Minneboo, A., Heymans, M. W., Lankhorst, G. J., Polman, C. H., & Bouter, L. M. (2009). Physical and Cognitive Functioning After 3 Years Can Be Predicted Using Information From the Diagnostic Process in Recently Diagnosed Multiple Sclerosis. *Archives of Physical Medicine and Rehabilitation*, 90(9), 1478-1488. <https://doi.org/10.1016/j.apmr.2009.03.018>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

E-mail address:

vuresearchportal.ub@vu.nl

Physical and Cognitive Functioning After 3 Years Can Be Predicted Using Information From the Diagnostic Process in Recently Diagnosed Multiple Sclerosis

Vincent de Groot, MD, PhD, Heleen Beckerman, PT, PhD, Bernard M. Uitdehaag, MD, PhD, Rogier Q. Hintzen, MD, PhD, Arjan Minneboo, MD, Martijn W. Heymans, PhD, Gustaaf J. Lankhorst, MD, Chris H. Polman, MD, Lex M. Bouter, PhD, on behalf of the Functional Prognostication and Disability (FuPro) Study Group

ABSTRACT. de Groot V, Beckerman H, Uitdehaag BM, Hintzen RQ, Minneboo A, Heymans MW, Lankhorst GJ, Polman CH, Bouter LM, on behalf of the Functional Prognostication and Disability (FuPro) Study Group. Physical and cognitive functioning after 3 years can be predicted using information from the diagnostic process in recently diagnosed multiple sclerosis. *Arch Phys Med Rehabil* 2009;90:1478-88.

Objective: To predict functioning after 3 years in patients with recently diagnosed multiple sclerosis (MS).

Design: Inception cohort with 3 years of follow-up. At baseline, predictors were obtained from medical history taking, neurologic examination, and magnetic resonance imaging (MRI).

Setting: Neurology outpatient clinic.

Participants: Patients with MS (N=156); 146 with complete follow-up.

Interventions: Not applicable.

Main Outcome Measures: Inability to walk at least 500m, impaired dexterity, cognitive impairments, incontinence, inability to drive a car or use public transportation, social dysfunction, and reliance on a disability pension.

Results: Clinical prediction rules were constructed for the models that were well calibrated (sufficient agreement between predicted and observed outcomes, based on visual inspection of calibration curves) and that showed sufficient discrimination (area under the receiver operation characteristic curve >.70) after internal bootstrap validation. The models for the inability to walk at least 500m, impaired dexterity, and cognitive impairments were well calibrated. Discrimination was sufficient for all 7 models, except the one predicting social dysfunction (.67). The inability to walk at least 500m was predicted by the

perceived ability to walk, impairment of the cerebellar tract, and the number of MRI lesions in the spinal cord. Impaired dexterity was predicted by the perceived ability to use the hands, impairments of the pyramidal, cerebellar, and sensory tracts, and the T2-weighted infratentorial lesion load. Cognitive impairment was predicted by age, gender, the perceived ability to concentrate, and the T2-weighted supratentorial lesion load.

Conclusions: Inability to walk at least 500m, impaired dexterity, and cognitive impairments can be predicted with predictors that are derived from medical history taking, neurologic examination, and MRI shortly after a definite diagnosis of MS has been made.

Key Words: Cohort studies; Disability evaluation; Multiple sclerosis; Prognosis; Rehabilitation.

© 2009 by the American Congress of Rehabilitation Medicine

MULTIPLE SCLEROSIS is characterized by variable neurologic symptomatology that differs not only between patients but also within patients over time. This variability makes predicting the clinical course of the disease difficult, posing a significant challenge for physicians treating patients with MS and causing patients to feel uncertain about their future. This uncertainty negatively influences their quality of life.^{1,2} Well-validated prognostic models can aid physicians in making decisions about certain (preventive) treatments for patients with MS or can improve the information given to these patients about their future prognosis.

Thus far, the prediction models published in the literature on MS have had a strong focus on the strength and the relevance of the predictors themselves,³⁻¹⁰ hoping that this would provide clues to a better understanding of the etiology or the course of the disease. Research that aims to investigate the strength of the relationship of a determinant with a particular outcome should focus on one determinant and correct for confounding variables in order to assess the real relationship between this determinant and the outcome. Reviews of the studies that have investigated determinants of the course of MS have shown that a progressive onset, being older at the time of diagnosis, an interval of less than 1 year between relapses, and impairments of pyramidal or cerebellar tracts are associated with a progressive disease

From the Departments of Rehabilitation Medicine (de Groot, Beckerman, Lankhorst), Clinical Epidemiology and Biostatistics (Uitdehaag), Neurology (Uitdehaag, Polman), and Radiology (Minneboo), and the EMGO Institute (de Groot, Beckerman, Heymans, Lankhorst, Bouter), VU University Medical Center, Amsterdam, The Netherlands; and Department of Neurology, Erasmus MC, Rotterdam, The Netherlands (Hintzen).

Supported by The Netherlands Organization for Scientific Research (grant no. NWO 940-33-009).

No commercial party having a direct financial interest in the results of the research supporting this article has or will confer a benefit upon the authors or upon any organization with which the authors are associated.

The Functional Prognostication and Disability (FuPro) Study Group includes the following investigators: G.J. Lankhorst, J. Dekker, A.J. Dallmeijer, M.J. IJzerman, H. Beckerman, V. de Groot: VU University Medical Center Amsterdam (project coordination); A.J.H. Prevo, E. Lindeman, V.P.M. Schepers: University Medical Center, Utrecht; H.J. Stam, E. Odding, B. van Baalen: Erasmus Medical Center, Rotterdam; A. Beelen, I.J.M. de Groot: Academic Medical Center, Amsterdam.

Reprint requests to Vincent de Groot, MD, PhD, Department of Rehabilitation Medicine, VU University Medical Center, PO Box 7057, 1007 MB Amsterdam, The Netherlands, e-mail: v.degroot@vumc.nl.

0003-9993/09/9009-0095\$36.00/0

doi:10.1016/j.apmr.2009.03.018

List of Abbreviations

AUC	area under the receiver operation characteristic curve
EDSS	Expanded Disability Status Scale
MRI	magnetic resonance imaging
MS	multiple sclerosis

course, whereas an exacerbation as a first sign of MS, a high recovery rate after the first exacerbation, and afferent or monoregional symptoms are associated with a more favorable disease course.³⁻¹⁰ This research has provided very useful information on the strength of the determinants themselves and has improved our understanding of the disease, but whether these determinants can be used to improve prognostication in individual patients has not been investigated.

In contrast with the literature on cardiac disorders,¹¹ intensive care units,¹² traumatic brain injury,¹³ and Guillain-Barré syndrome,¹⁴ the literature on MS has not yet assessed the usefulness of the complete prognostic models to predict future events accurately. The construction of a complete prognostic model differs fundamentally from research that investigates the strength of a determinant.¹⁵ All phases of the development of a prognostic model are directed towards obtaining a model that maintains its prognostic ability in different clinical samples of patients. This means that determinants that are easily obtainable in clinical practice are preferred above highly specialized measurements that are not routinely collected, and that predictors that are already known from the medical literature and (expert) clinicians will be used for the model construction. Furthermore, during the construction of the regression models, less emphasis is placed on the significance level of the determinants, which often means that a (very) liberal *P* value is used. With this strategy, the risk of overfitting the regression models is minimized, and the chance of obtaining an externally valid model is increased. Finally, in the presentation of the results, the accuracy of the predictions of the whole model is emphasized. For MS, one prognostic study¹⁶ to assess the risk of reaching secondary progression has been published, but this study used a different approach, namely a Bayesian analysis, to assess the risk. In a large sample, the risk of several determinants, which were selected on the basis of a previous study, was calculated. The specificity of the model was very good, while the sensitivity was poor.

With respect to future outcomes, most studies have focused on neurologic and locomotor function, using the score of the EDSS as the outcome and the neurologic deficits or MRI parameters as candidate predictors. However, other areas of functioning are relevant for patients, such as wheelchair dependence, impaired dexterity, cognitive impairments, incontinence, inability to use a car or public transportation, social dysfunction, and reliance on a disability pension. Studying these outcomes also means that the predictors should not be limited to neurologic or MRI parameters, but that psychosocial predictors should also be assessed.

The aim of our study was to construct and assess the usefulness of prediction models to predict functioning in the areas of mobility, dexterity, cognition, voiding, transportation, social activities, and work.

METHODS

Patients and Design

All consecutive, potentially eligible patients visiting the participating outpatient clinics of 5 neurology departments were invited to participate. A cohort of 156 patients, aged 16 to 55 years, with recently (<6mo previously) diagnosed MS was recruited from 1998 to 2000 and prospectively monitored for 3 years. Diagnosis was determined according to the Poser criteria for definite MS.¹⁷ Treatments were not standardized. Patients with other neurologic disorders, systemic diseases, or malignant neoplastic diseases were excluded. This study was performed as part of a longitudinal study collecting extensive data on many potentially relevant predictors and outcomes at base-

line and at 6 months, 1, 2, and 3 years later.^{18,19} For the present analyses we used the baseline information for the predictors, and the 3-year data for the relevant outcomes. The patients were visited at home to minimize dropout, and 4 well-trained raters were responsible for the scoring. The ethics committee of the VU University Medical Center approved this study.

Construction of Prediction Models

As has been outlined in the introduction, the construction of a prediction model requires a specific methodological approach.¹⁵ The prediction models were constructed with the intention to use them in clinical practice. Therefore, we involved representatives of potential users of these models in the construction phase. Before actual data analysis, the aims of our study were discussed during 2 informal semistructured workshops with neurologists and researchers specializing in MS, and with rehabilitation physicians and physical and occupational therapists. In these workshops, we discussed which outcomes would be relevant to predict, and which candidate predictors should be investigated to predict these outcomes.

Outcomes

Inability to walk at least 500m was defined as an EDSS score of 4 or higher.²⁰ Impaired dexterity was defined as an abnormal score (mean – 1.96 SD, healthy Dutch reference population) for the 9-Hole Peg Test.²¹ Cognitive impairments were defined as a score of mean – SD for 1 or more subtests of a cognitive screening test that was specifically developed for MS, which includes the subscales Consistent Long Term Retrieval and Long Term Storage of the Selective Reminding Test measuring verbal learning and memory, the 10/36 Spatial Recall Test measuring visuospatial learning and delayed recall, the Symbol Digit Modalities Test measuring sustained attention and concentration, the Paced Auditory Serial Addition Test measuring sustained attention and information processing speed, and the Word List Generation measuring verbal fluency.²²⁻²⁴ Incontinence was defined as a score of 5 or lower for the continence item of the FIM.²⁵ Inability to drive a car or use public transportation was defined as needing help or being unable on the ability to travel item of the Rehabilitation Activities Profile.²⁶ Social dysfunction was defined as an abnormal score (mean – 1.96 SD, healthy Dutch reference population) for 1 or more of the 3 social subscales (role physical, role emotional, social functioning) of the Medical Outcomes Study 36-Item Short-Form Health Survey.²⁷ The patients were asked in a direct question about complete or partial reliance on a disability pension.

Candidate Predictors

Participants in the workshops were encouraged to name predictors that are relatively easy to obtain in clinical practice. First, the most relevant predictors for which information could be gathered during medical history taking were identified. Next, the most relevant predictors for which a physical examination is required were identified, and finally, the most relevant predictors obtained through complex diagnostic tests were identified. Using the information obtained from the discussions and from the literature, as described in the introduction, we selected candidate predictors from the baseline data of the extensive data set.¹⁹ Table 1 shows the selected outcomes and the predictors that were used to construct the models. Data on the selected outcomes obtained at baseline were not used as predictors. For the predictors that are based on medical history taking we used items of the Disability and Impact Profile.^{28,29} This written questionnaire contains patient-rated numerical rating scales, which range from 0 to 10, on 40 different abilities.

Table 1: Candidate Predictors Measured at Baseline for Each Outcome of Interest

Predictor per Outcome of Interest	Range	Description
Inability to walk at least 500m		
Medical history taking		
• How well can you walk?*	0–10	Not at all—very well
• Are you easily tired?*	0–10	Very easily—not at all
Physical examination		
• Impairment of pyramidal tract [†]	0–6	No signs—quadriplegia
• Impairment of cerebellar tract [†]	0–5	No signs—severe ataxia
MRI-parameter		
• No. of lesions in spinal cord [‡]	n	No. of lesions counted
Impaired dexterity		
Medical history taking		
• How well can you use your hands?*	0–10	Not at all—very well
Physical examination		
• Impairment of sensory tract [†]	0–6	No signs—sensation lost below head
• Impairment of pyramidal tract [†]	0–6	No signs—quadriplegia
• Impairment of cerebellar tract [†]	0–5	No signs—severe ataxia
MRI-parameter		
• T2-weighted infratentorial lesion load [‡]	cm ³	
Cognitive impairments		
Medical history taking		
• Age	y	
• Gender	0–1	Woman—man
• How good is your memory?*	0–10	Bad—good
• How well can you concentrate?*	0–10	Not at all—very well
Physical examination		
• None		
MRI-parameter		
• T2-weighted supratentorial lesion load [‡]	cm ³	
Incontinence		
Medical history taking		
• Can you contain your urine well?*	0–10	Not at all—easily
Physical examination		
• Impairment of pyramidal tract [†]	0–6	No signs—quadriplegia
MRI-parameter		
• No. of lesions in spinal cord [‡]	n	No. of lesions counted
Inability to use a car or public transportation		
Medical history taking		
• How good is your memory?*	0–10	Bad—good
• How well can you concentrate?*	0–10	Not at all—very well
Physical examination		
• Impairment of pyramidal tract [†]	0–6	No signs—quadriplegia
• Impairment of cerebellar tract [†]	0–5	No signs—severe ataxia
MRI-parameter		
• T2-weighted total lesion load [‡]	cm ³	
• No. of lesions in spinal cord [‡]	n	No. of lesions counted
Social dysfunction		
Medical history taking		
• How good is your contact with members of your household?*	0–10	Bad—excellent
• How do you feel?*	0–10	Gloomy—happy
• Are you easily tired?*	0–10	Very easily—not at all
Physical examination		
• Impairment of pyramidal tract [†]	0–6	No signs—quadriplegia
• Impairment of cerebellar tract [†]	0–5	No signs—severe ataxia
MRI-parameter		
• T2-weighted total lesion load [‡]	cm ³	
Reliance on a disability pension		
Medical history taking		
• How do you feel?*	0–10	Gloomy—happy
• How good is your memory?*	0–10	Bad—good
• How well can you concentrate?*	0–10	Not at all—very well
• Are you easily tired?*	0–10	Very easily—not at all

Table 1 (Cont'd): Candidate Predictors Measured at Baseline for Each Outcome of Interest

Predictor per Outcome of Interest	Range	Description
Physical examination		
• Impairment of pyramidal tract [†]	0–6	No signs—quadriplegia
• Impairment of cerebellar tract [†]	0–5	No signs—severe ataxia
MRI-parameter		
• T2-weighted total lesion load [‡]	cm ³	

*Item of the Disability and Impact Profile.

[†]Item of the Functional Systems of the EDSS.

[‡]Values derived from MRI of the brain and spinal cord.

Each ability is assessed with 2 questions: (1) a question to assess the perceived disability for that item, and (2) a question to assess the extent to which the perceived disability poses a problem for the patient. We used the first question of the abilities that we were interested in. For the predictors that are based on physical examination, we used the EDSS Functional Systems scores.²⁰ MRI was used to obtain the predictor variables T2-weighted (supra- and infratentorial) lesion loads in cm³, and the number of lesions in the spinal cord.^{30,31} In total, 4 MRI predictor variables were used: T2 supratentorial, T2 infratentorial, T2 total, and spinal cord.

Analysis

Only patients with complete outcome data at 3 years were analyzed. To improve data quality and reduce the risk of bias, missing data on predictors were imputed twice^{32,33} by using the data augmentation procedure in NORM software,³⁴ yielding 2 imputed data sets. Descriptive statistics were used to describe the study population. For each outcome the number and percentage of patients with an unfavorable outcome were calculated.

Because predictive modeling in small data sets is susceptible to bias, we made use of the approach described by Steyerberg et al,^{15,32,33} which we described in the introduction. We used a limited set of candidate predictors that were selected on the basis of information from the literature and on clinical grounds. Subsequently, logistic regression models were constructed in each imputed data set, using a backwards stepwise selection procedure with a liberal *P* value of 0.5. When predictors in these models showed a counterintuitive relationship with the outcome, which means that the sign of the regression coefficient is opposite to what we expected, this predictor was deleted from the model, and the backwards selection procedure was repeated. Because the selected predictors were the same in both imputed data sets, internal validation was performed on one of the sets.

Bootstrapping techniques were used to study the internal validity of the final models (ie, to adjust the estimated regression coefficients for overfitting and the model performance for overoptimism).^{33,35} Random bootstrap samples were drawn with replacement (250 replications) from the full data set. The shrinkage factor, a result of the bootstrap analyses, is a measure of overfitting. Regression coefficients can be corrected for overfitting by multiplying them by this shrinkage factor. Bootstrapping was performed in S-plus 6.1.^a

Model Performance

The model performance, expressed as calibration and discrimination, after bootstrapping can be considered as the performance that can be expected from similar future patients. Calibration refers to whether the predicted outcomes agree with the observed outcomes. A frequently occurring problem with prediction models is that the predictions for new patients are

too extreme (too high for high-risk patients and too low for low-risk patients). Well-calibrated models have a slope of 1, while models providing predictions that are too extreme have a slope of less than 1.

The discriminative ability of the model (ie, how accurately can high-risk patients be distinguished from low-risk patients) was assessed using the AUC (95% confidence interval). An AUC of 0.5 indicates no discrimination above chance, whereas an AUC of 1.0 indicates perfect discrimination. A rough guide for classifying the discriminative ability of a diagnostic test is the traditional academic points system: excellent (>.90), good (>.80), fair (>.70), poor (>.60), or fail (>.50).³⁶

Clinical Prediction Rules

To facilitate the calculation of an individual patient's risk, we developed score charts for the prediction models that were internally valid. We divided the regression coefficients of the multivariate models by the lowest regression coefficient and rounded them to the nearest integer to form scores for the predictors. The sum of the scores corresponds to the risk of a poor outcome. We created 3 risk categories: high (probability of adverse outcome >75%), moderate (probability of adverse outcome 25%–75%), and low (probability of adverse outcome <25%).

Table 2: Baseline Characteristics (n=146)

Patient characteristics	
Women, n (%)	93 (64)
Age (mean ± SD) (y)	37.4±9.7
Disease characteristics	
Relapse (RO) vs nonrelapse (NRO) onset	82% RO
EDSS	2.5 (2.0–3.0)
Candidate predictors	
How well can you walk?	9 (7–10)
How well can you use your hands?	9 (8–10)
Can you contain your urine well?	9 (7–10)
How good is your contact with members of your household?	10 (8–10)
How good is your memory?	8 (7–9)
How well can you concentrate?	8 (7–9)
How do you feel?	8 (6–8)
Are you quickly fatigued?	7 (6–9)
Impairment of sensory tract	1 (1–2)
Impairment of pyramidal tract	1 (0–1)
Impairment of cerebellar tract	1 (0–2)
T2-weighted supratentorial lesion load (cm ³)	3.4 (0.8–11.3)
T2-weighted infratentorial lesion load (cm ³)	0.2 (0–0.5)
T2-weighted total lesion load (cm ³)	3.6 (1–11.4)
No. of lesions in spinal cord	2 (1–4)

NOTE. Values are median (interquartile range) unless otherwise indicated.

Table 3: Frequencies of Unfavorable Outcomes at Baseline and After 3 Years (n=146)

	Baseline	Changes		3y
		Improved	Deteriorated	
Inability to walk at least 500m	16 (11.0)	5 (3.4)	26 (17.8)	37 (25.3)
Impaired dexterity	36 (24.7)	4 (2.7)	14 (9.6)	46 (31.5)
Cognitive impairments	60 (41.1)	29 (19.9)	13 (8.9)	44 (30.1)
Incontinence	9 (6.2)	6 (4.1)	21 (14.4)	24 (16.4)
Inability to use a car or public transportation	9 (6.2)	6 (4.1)	11 (7.5)	14 (9.6)
Social dysfunction	58 (39.7)	20 (13.7)	22 (15.1)	60 (41.1)
Reliance on a disability pension	26 (17.8)	3 (2.1)	54 (37.0)	77 (52.7)

NOTE. Values are n (%).

RESULTS

Patients

Data on the outcomes at the 3-year follow-up were missing for 10 of the 156 patients. These 10 patients did not differ significantly from the rest of the cohort with regard to gender, age, T2-weighted lesion load at baseline, or number of lesions in the spinal cord at baseline. However, they had a trend towards higher baseline EDSS scores and, in contrast to the

results for the EDSS, fewer lesions on the baseline MRI. For 13 of the 146 patients with a complete follow-up, baseline MRI data on the brain and spinal cord were missing. MRI data on the spinal cord were also missing for 2 patients. These data were imputed. Data on all other candidate predictors were complete. Table 2 shows the baseline characteristics of the patients, most of which are consistent with the expected pattern: more women than men, and approximately 80% with a relapse onset.

Table 4: Final Regression Models and Their Predictive Ability

Models and Predictors (Score Range)	Predictive Value			Model Performance	
	β_{shrunk}	Factor	P	Slope	AUC (95% CI)
Inability to walk at least 500m					
How well can you walk (0–10)?	-.57	3	.00	.93	.89 (.83–.95)
Impairment cerebellar tract (0–5)	.77	-5	.00		
No. of lesions in spinal cord	.16	-1	.05		
Impaired dexterity					
How well can you use your hands (0–10)?	-.16	1	.16	.85	.77 (.69–.86)
Impairment pyramidal tract (0–6)	.25	-2	.31		
Impairment cerebellar tract (0–5)	.46	-3	.03		
Impairment sensory tract (0–6)	.27	-2	.17		
T2-weighted infratentorial lesion load	.97	-6	.00		
Cognitive impairments					
Age	.03	1	.12	.88	.74 (.65–.83)
Gender	.88	29	.02		
How well can you concentrate (0–10)?	-.17	-5	.07		
T2-weighted supratentorial lesion load	.06	2	.00		
Incontinence					
Can you contain your urine well (0–10)?	-.44		.00	.97	.80 (.71–.90)
No. of lesions in spinal cord	.10		.25		
Inability to use a car or public transportation					
How good is your memory (0–10)?	-.19		.06	.71	.76 (.65–.87)
Impairment of pyramidal tract (0–6)	.38		.20		
Impairment of cerebellar tract (0–5)	.29		.25		
No. of lesions in spinal cord	.12		.09		
Social dysfunction					
How good is your contact with members of your household (0–10)?	-.20		.06	.87	.67 (.58–.76)
Are you easily tired (0–10)?	.16		.01		
T2-weighted total lesion load	.01		.27		
Reliance on a disability pension					
How well can you concentrate (0–10)?	-.23		.01	.84	.72 (.64–.80)
Impairment of pyramidal tract (0–6)	.17		.44		
Impairment of cerebellar tract (0–5)	.19		.29		
T2-weighted total lesion load	.03		.08		

NOTE. Results for final models after internal validation by means of 250 bootstraps. $\beta_{\text{shrunk}} = \beta_{\text{original}} \times \text{slope}$. Factor = $\beta_{\text{shrunk}} / \text{lowest } \beta_{\text{shrunk}}$ of the multivariate model, rounded to the nearest integer for use in the score charts. Slope: shrinkage factor obtained after bootstrapping, well-calibrated models have a slope of 1. AUC: .50 indicates no discrimination beyond chance, >.70 indicates sufficient discrimination. Abbreviation: CI, confidence interval.

Table 3 shows the number of patients with an unfavorable outcome at baseline and at the 3-year follow-up. For most patients, functioning does not change over the 3-year period. Most changes are in the direction of unfavorable outcomes. Exceptions are the outcomes of cognitive impairment (29 patients showed remarkable improvement) and social functioning (important changes in both directions).

The final regression models, obtained after a backwards stepwise procedure with a liberal P value of 0.5 and after elimination of predictors with a counterintuitive relationship with the outcome, are shown in table 4. The presented models are corrected for overoptimism by bootstrapping. Figure 1 shows the discrimination and calibration curves. The outcomes for inability to walk at least 500m, impaired dexterity, and cognitive impairments show good calibration (calibration curves follow approximately the 45° diagonal, and the shrinkage factors [slope] approach 1). The calibration curves for the other outcomes show important miscalibration. Discriminative ability is good for the models predicting inability to walk at least 500m (AUC=.89 [.83–.95]) and incontinence (AUC=.80 [.71–.90]); fair for the models predicting impaired dexterity (AUC=.77 [.69–.86]), cognitive impairments (AUC=.74 [.65–.83]), inability to use a car or public transportation (AUC=.76 [.65–.87]), and reliance on a disability pension (AUC=.72 [.64–.80]); and poor for the model predicting social dysfunction (AUC=.67 [.58–.76]).

Table 4 also shows that information obtained from medical history taking and MRI is included in every regression model, and that information obtained from the physical examination is not included in the models that predict incontinence and social dysfunction.

Twelve of the 37 potential predictors did not predict the outcome they were supposed to predict. Seven omitted predictors were from the category medical history taking (“Are you easily tired?” [2x], “How good is your memory?” [2x], “How well can you concentrate?”, “How do you feel?” [2x]); 4 from the category physical examination (impairment pyramidal [3x] and cerebellar tracts); and 1 was an MRI parameter (T2-weighted lesion load). However, of these, only “How do you feel?” did not predict any outcome it was supposed to predict.

Clinical Prediction Rules

Clinical prediction rules were constructed for the models predicting inability to walk at least 500m, impaired dexterity, and cognitive impairments (appendix 1). They are fully based on the results of the final regression models. The “factors” from table 4 are used in the calculations of the clinical prediction rules.

DISCUSSION

We have shown that it is feasible to make internally valid predictions for patients with recently diagnosed MS with regard to outcomes on physical and cognitive functioning. The inability to walk at least 500m was predicted by the perceived ability to walk, impairment of the cerebellar tract, and the number of MRI lesions in the spinal cord. Impaired dexterity was predicted by the perceived ability to use the hands, impairments of the pyramidal, cerebellar, and sensory tracts, and the T2-weighted infratentorial lesion load. Cognitive impairments were predicted by age, gender, the perceived ability to concentrate, and the T2-weighted supratentorial lesion load.

In general, our results show that it makes sense to select potential predictors by following the diagnostic process of the physician (ie, first, medical history taking; then physical examination; finally MRI), because all prediction models contain information from medical history taking and MRI, and only 2 of the 7

prediction models did unexpectedly not contain predictors from the physical examination. Similarly, Bergamaschi et al¹⁶ suggested incorporating additional clinical information, such as information on fatigue, cognitive impairments, and neuroradiologic information, into their prediction model in order to improve the sensitivity. In addition, they also suggested incorporating genetic, neuroimmunologic, and neurophysiologic information. Although impairments of the pyramidal tract are frequently accompanied by bladder problems, apparently they do not contribute to the prediction of incontinence. Also, we wrongly expected impairments of the pyramidal and cerebellar tracts to predict social functioning. Nevertheless, we think that our results show that useful prognostic information can be obtained from the standard routine of information gathering in clinical practice.

It is very tempting to (causally) interpret the strength of the associations between the predictors in the final models and the predicted outcomes. However, as outlined in the introduction, we have used a specific method to construct the regression models. The aim of this method is to predict future events as accurately as possible and not to assess the strength of an association. Most importantly, this method does not investigate confounding, which means that an assessment of the unconfounded association is not possible, and thus interpreting results in this way should not be done. In contrast to the method that we describe in this article, we have published an article¹⁹ in which we used a completely different method of analyzing our longitudinal data with the intention to identify the most powerful determinants of social functioning. It is also very tempting to add other clinical, or new, potentially stronger determinants to these models. An example may be brain atrophy measurements in the model for cognitive functioning. Although brain atrophy has been suggested to be causally related to cognitive functioning, adding this information to a prediction model does not necessarily mean that predictions improve. In prediction modeling, the added value of a determinant should be investigated by assessing the change in discriminative ability (AUC) and model fit, and not by looking at the strength of the association.

An important strength of our study is that the analysis was designed to optimize the internal validity.^{32,33} Several attempts were made to minimize bias. First, missing baseline data were imputed to optimize the quality of the data. Second, we used a limited set of clinically relevant candidate predictors that were only excluded when the P value was greater than .50, or when the sign of the coefficient was opposite to what we expected. Finally, bootstrapping was used to correct for overoptimism of the regression coefficients and the model parameters (calibration: shrinkage factor, and discrimination: AUC).

Study Limitations

A possible weakness of the study was the assessment of cognitive dysfunction. Twenty-nine patients showed cognitive improvements in the first 3 years, substantially more than the number of patients who improved on the other outcomes. In accordance with the design of our study,¹⁸ cognitive data were collected annually, but it is possible that an interval of 1 year is not sufficiently long to rule out a practice effect. Another explanation might be that the definition of cognitive impairment that we applied does not correctly diagnose cognitive impairment in patients. The cognitive screening test is based on 5 cognitive tests that each assess a different aspect of cognitive functioning, but in the literature there is no consensus on which cutoff point to use.^{23,24,37–39} We used a sensitive cutoff point that classified patients as cognitively impaired if 1 or more of their test scores were lower than the mean – SD, compared with a Dutch reference population. Our strategy might therefore lead to a greater number of patients classified as cognitively im-

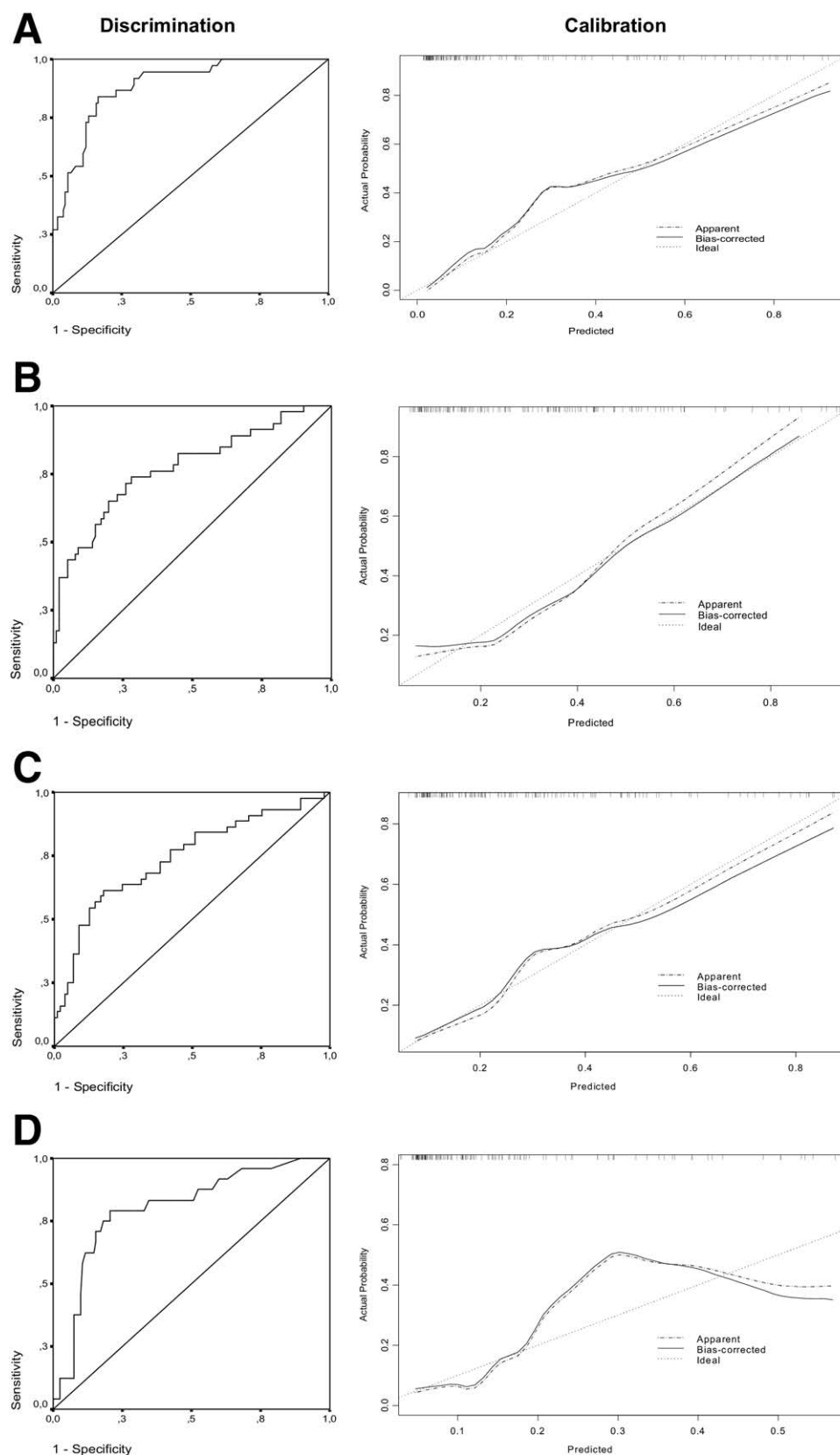


Fig 1. Discrimination (left column) and calibration (right column) curves for all outcomes. The ideal line represents perfect calibration, the apparent line represents our original data and the bias-corrected line represents the bootstrap corrected calibration of the model. (A) Inability to walk at least 500 meters, (B) impaired dexterity, (C) cognitive impairments, (D) incontinence.

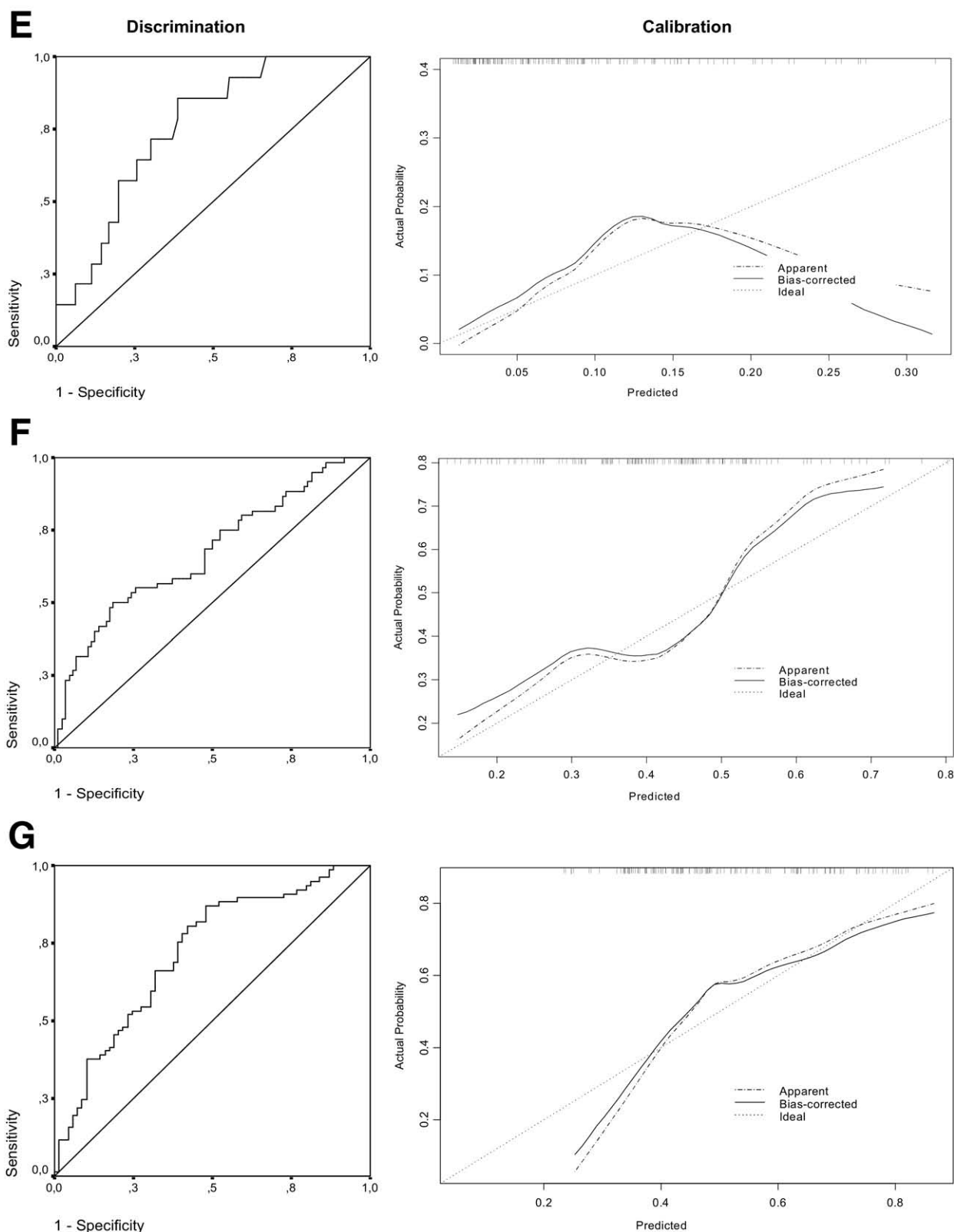


Fig 1. (E) inability to use a car or public transportation, (F) social dysfunction, (G) reliance on a disability pension.

paired, whereas they actually perform within the norm (ie, patients are classified as false-positives). Therefore, the observed improvements in cognitive functioning might just be changes that occur within normal ranges. Alternative cutoff points, such as 2 or more test scores lower than the mean – SD, or 1 or more test scores lower than the mean – 2SD, have also been applied in the literature. However, applying these criteria to our data still showed cognitive improvements for a substantial number of patients (data not shown). Therefore, the observed improvements in cognitive functioning are either caused by a practice effect or they are real improvements.

At baseline (ie, a maximum of 6 months after a definite diagnosis of MS was made) 9 (6%) patients were receiving disease-modifying treatment. At the 3-year follow-up, this rose to 44 (30%) with a mean treatment duration of 25 months. We did not include disease-modifying treatment at baseline in our models because we assumed that confounding by indication could influence our findings. Patients with a more severe disease course are more likely to receive this treatment. The omission of disease-modifying treatment in the prediction models means that our models can be used independent of disease-modifying treatment. With regard to external validity, this means that our results can be generalized to populations in which approximately the same percentage of patients are receiving disease-modifying treatment.

Although our results look promising, application in clinical practice is not justified until they have been validated externally.⁴⁰⁻⁴³ The analyses that we have presented should be repeated in a new cohort, which should be recruited in a different geographic area, at a different point in time, or, as is current in MS, assessed with different diagnostic criteria.⁴⁴ The regression coefficients and model parameters in these cohorts should be used to assess the applicability of these models in clinical practice. When external validation has shown that the models perform well, and when the clinical usefulness of the clinical prediction rules has been established, they can be used with confidence in clinical practice to aid clinicians in making a prognosis. However, because the application of research findings in clinical practice is not self-evident, the clinical prediction rules should be actively implemented.^{45,46}

Our results indicate that predictions of the outcomes that are based on performance measures (ie, measures that require patients to actually perform a physical or cognitive test) are better than the predictions of outcomes based on self-reported health status. This implies that the more objective outcomes can be correctly predicted, but that self-reported outcomes are more difficult to predict. The reason for this might be that personal or social factors, which are not easy to measure as predictors, also have an effect on self-reported outcomes. In clinical practice, the clinical prediction rules could be used not only to improve treatment decisions regarding the initiation of disease-modifying treatment, but also to improve the timing of the (components of) rehabilitation treatment. Of equal importance is the possibility to improve the counseling of a patient. In conversations with the patient, the physician should become familiar with the patient's personal and social situation. When this information is combined with the information obtained from the clinical prediction rules, a patient-specific prognosis can be formulated, which the physician can then discuss with the patient. The results of this discussion can be used to adjust the counseling of the patient, or can lead to the initiation of preventive measures or (rehabilitation) treatment.

CONCLUSIONS

In conclusion, during the first 3 years of MS, it is possible to predict accurately inability to walk at least 500m, impaired dexterity, and cognitive impairments based on predictors that are derived from medical history taking, physical examination, and MRI shortly after a definite diagnosis of MS has been made. The ability to predict physical and cognitive functioning might facilitate the counseling of patients and the planning of (rehabilitation) treatment. But first, adequate performance of the models in a new cohort must be validated externally.

Acknowledgments: We thank the neurologists in the participating hospitals (VU University Medical Center, Academic Medical Center Amsterdam, Sint Lucas Andreas Hospital Amsterdam, OLVG Hospital Amsterdam, Erasmus Medical Center Rotterdam) for recruiting the patients, and M. Jacobs-Van der Bruggen, PT, M. Schothorst, PT, and T. Wedding, PT, for performing the measurements.

APPENDIX 1: CLINICAL PREDICTION RULES

Clinical Prediction Rules									
Inability to walk at least 500m									
How well can you walk (0–10)?	_____	x	3	=	_____			Risk	Score
Impairment of cerebellar tract (0–5)	_____	x	–5	=	_____			High	<13
No. of lesions in spinal cord	_____	x	–1	=	_____	+		Moderate	13–20
								Low	>20
			Score	=	<input type="text"/>				
Impaired dexterity									
How well can you use your hands (0–10)?	_____	x	1	=	_____			Risk	Score
Impairment of pyramidal tract (0–6)	_____	x	–2	=	_____			High	<–5
Impairment of cerebellar tract (0–5)	_____	x	–3	=	_____			Moderate	2 to –5
Impairment of sensory tract (0–6)	_____	x	–2	=	_____			Low	>2
T2-weighted infratentorial lesion load (cm ³)	_____	x	–6	=	_____	+			
			Score	=	<input type="text"/>				
Cognitive impairments									
Age	_____	x	1	=	_____			Risk	Score
Man	_____	+	29	=	_____			High	>40
How well can you concentrate (0–10)?	_____	x	–5	=	_____			Moderate	3–40
T2-weighted supratentorial lesion load (cm ³)	_____	x	2	=	_____	+		Low	<3
			Score	=	<input type="text"/>				

APPENDIX 1: CLINICAL PREDICTION RULES (Cont'd)

Risk: High = probability of adverse outcome >75%; Moderate = probability of adverse outcome 25%–75%; Low = probability of adverse outcome <25%.

Impairment of cerebellar tract

- 0 = Normal
- 1 = Abnormal signs without disability
- 2 = Mild ataxia
- 3 = Moderate truncal or limb ataxia
- 4 = Severe ataxia
- 5 = Unable to perform coordinated movements

Impairment of pyramidal tract

- 0 = Normal
- 1 = Abnormal signs without disability
- 2 = Monoparesis grade 4
- 3 = Monoparesis grade 2/3, or paraparesis/hemiparesis grade 3/4
- 4 = Monoparesis grade 0/1, or paraparesis/hemiparesis grade 2, or tetraparesis grade 3/4
- 5 = Paraparesis/hemiparesis grade 0/1, or tetraparesis grade 2
- 6 = Tetraparesis grade 0/1

Impairment of sensory tract

- 0 = Normal
- 1 = Vibration/figure-writing mildly decreased 1–2 limbs
- 2 = Vibration mildly decreased 3–4 limbs, or vibration moderately decreased 1–2 limbs, or touch/pain mildly decreased 1–2 limbs
- 3 = Vibration lost 1–2 limbs, or touch/pain moderately decreased 1–2 limbs, or proprioception moderately decreased 3–4 limbs
- 4 = Touch/pain moderately decreased 3–4 limbs, or touch/pain markedly decreased 1–2 limbs, or proprioception markedly decreased 3–4 limbs, or proprioception lost 1–2 limbs
- 5 = Touch/pain moderately decreased below head, or sensation lost 1–2 limbs, or proprioception lost below the head
- 6 = Sensation lost below the head

References

1. Janssens AC, De Boer JB, Van Doorn PA, et al. Expectations of wheelchair-dependency in recently diagnosed patients with multiple sclerosis and their partners. *Eur J Neurol* 2003;10:287-93.
2. Janssens AC, Van Doorn PA, De Boer JB, Van der Meché FG, Passchier J, Hintzen RQ. Perception of prognostic risk in patients with multiple sclerosis: the relationship with anxiety, depression, and disease-related distress. *J Clin Epidemiol* 2004;57:180-6.
3. Runmarker B, Andersson C, Oden A, Andersen O. Prediction of outcome in multiple sclerosis based on multivariate models. *J Neurol* 1994;241:597-604.
4. Runmarker B, Andersen O. Prognostic factors in a multiple sclerosis incidence cohort with twenty-five years of follow-up. *Brain* 1993;116:117-34.
5. Weinshenker BG, Bass B, Rice GPA, et al. The natural history of multiple sclerosis: a geographically based study. 2. Predictive value of the early clinical course. *Brain* 1989;112:1419-28.
6. Weinshenker BG, Bass B, Rice GPA, et al. The natural history of multiple sclerosis: a geographically based study. I. Clinical course and disability. *Brain* 1989;112:133-46.
7. Weinshenker BG, Rice GPA, Noseworthy JH, Carriere W, Baskerville J, Ebers GC. The natural history of multiple sclerosis: a geographically based study. 3. Multivariate analysis of predictive factors and models of outcome. *Brain* 1991;114:1045-56.
8. Amato MP, Ponziani G. A prospective study on the prognosis of multiple sclerosis. *Neurol Sci* 2000;21:S831-8.
9. Confavreux C, Vukusic S, Adeleine P. Early clinical predictors and progression of irreversible disability in multiple sclerosis: an amnesic process. *Brain* 2003;126:770-82.
10. Pittock SJ, Mayr WT, McClelland RL, et al. Change in MS-related disability in a population-based cohort: a 10-year follow-up study. *Neurology* 2004;62:51-9.
11. Matheny ME, Ohno-Machado L, Resnic FS. Discrimination and calibration of mortality risk prediction models in interventional cardiology. *J Biomed Inform* 2005;38:367-75.
12. Harrison DA, Brady AR, Parry GJ, Carpenter JR, Rowan K. Recalibration of risk prediction models in a large multicenter cohort of admissions to adult, general critical care units in the United Kingdom. *Crit Care Med* 2006;34:1378-88.
13. Perel P, Arango M, Clayton T, et al. Predicting outcome after traumatic brain injury: practical prognostic models based on large cohort of international patients. *BMJ* 2008;336:425-9.
14. Van Koningsveld R, Steyerberg EW, Hughes RA, Swan AV, Van Doorn PA, Jacobs BC. A clinical prognostic scoring system for Guillain-Barre syndrome. *Lancet Neurol* 2007;6:589-94.
15. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med* 2000;19:1059-79.
16. Bergamaschi R, Quaglini S, Trojano M, et al. Early prediction of the long term evolution of multiple sclerosis: the Bayesian Risk Estimate for Multiple Sclerosis (BREMS) score. *J Neurol Neurosurg Psychiatry* 2007;78:757-9.
17. Poser CM, Paty DW, Scheinberg L, et al. New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Ann Neurol* 1983;13:227-31.
18. De Groot V, Beckerman H, Lankhorst GJ, Polman CH, Bouter LM. The initial course of daily functioning in multiple sclerosis: a three-year follow-up study. *Mult Scler* 2005;11:713-8.
19. De Groot V, Beckerman H, Twisk JW, et al. Vitality, perceived social support and disease activity determine the performance of social roles in recently diagnosed multiple sclerosis: a longitudinal analysis. *J Rehabil Med* 2008;40:151-7.
20. Kurtzke JF. Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS). *Neurology* 1983;33:1444-52.
21. Uitdehaag BMJ, Adèr HJ, Roosma TJ, De Groot V, Kalkers NF, Polman CH. Multiple sclerosis functional composite: impact of

- reference population and interpretation of changes. *Mult Scler* 2002;8:366-71.
22. Rao SM, Leo GJ, Bernardin L, Unverzagt F. Cognitive dysfunction in multiple sclerosis. I. Frequency, patterns, and prediction. *Neurology* 1991;41:685-91.
 23. Boringa JB, Lazeron RH, Reuling IE, et al. The brief repeatable battery of neuropsychological tests: normative values allow application in multiple sclerosis clinical practice. *Mult Scler* 2001;7:263-7.
 24. Achiron A, Barak Y. Cognitive impairment in probable multiple sclerosis. *J Neurol Neurosurg Psychiatry* 2003;74:443-6.
 25. Granger CV, Cotter AC, Hamilton BB, Fiedler RC, Hens MM. Functional assessment scales: a study of persons with multiple sclerosis. *Arch Phys Med Rehabil* 1990;71:870-5.
 26. Van Bennekom CAM, Jelles F, Lankhorst GJ, Bouter LM. The rehabilitation activities profile: a validation study of its use as a disability index with stroke patients. *Arch Phys Med Rehabil* 1995;76:501-7.
 27. Aaronson NK, Muller M, Cohen PD, et al. Translation, validation, and norming of the Dutch language version of the SF-36 Health Survey in community and chronic disease populations. *J Clin Epidemiol* 1998;51:1055-68.
 28. Laman H, Lankhorst GJ. Subjective weighting of disability: an approach to quality of life assessment in rehabilitation. *Disabil Rehabil* 1994;16:198-204.
 29. Lankhorst GJ, Jelles F, Smits RCF, et al. Quality of life in multiple sclerosis: the Disability and Impact Profile (DIP). *J Neurol* 1996;243:469-74.
 30. Kalkers NF, Bergers L, De Groot V, et al. Concurrent validity of the MS Functional Composite using MRI as a biological disease marker. *Neurology* 2001;56:215-9.
 31. Bot JC, Barkhof F, Polman CH, et al. Spinal cord abnormalities in recently diagnosed MS patients: added value of spinal MRI examination. *Neurology* 2004;62:226-33.
 32. Steyerberg EW, Eijkemans MJ, Habbema JD. Stepwise selection in small data sets: a simulation study of bias in logistic regression analysis. *J Clin Epidemiol* 1999;52:935-42.
 33. Steyerberg EW, Harrell FE Jr, Borsboom GJ, Eijkemans MJ, Vergouwe Y, Habbema JD. Internal validation of predictive models: efficiency of some procedures for logistic regression analysis. *J Clin Epidemiol* 2001;54:774-81.
 34. NORM: Multiple imputation of incomplete multivariate data under a normal model, version 2, 1999. Software for Windows 95/98/NT. Available at: <http://www.stat.psu.edu/~jls/misoftwa.html>. Accessed August 16, 2009.
 35. Harrell FE, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996;15:361-87.
 36. Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285-93.
 37. Solari A, Mancuso L, Motta A, Mendozzi L, Serrati C. Comparison of two brief neuropsychological batteries in people with multiple sclerosis. *Mult Scler* 2002;8:169-76.
 38. Aupperle RL, Beatty WW, Shelton FN, Gontkovsky ST. Three screening batteries to detect cognitive impairment in multiple sclerosis. *Mult Scler* 2002;8:382-9.
 39. Dent A, Lincoln NB. Screening for memory problems in multiple sclerosis. *Br J Clin Psychol* 2000;39:311-5.
 40. Hukkelhoven CW, Rampen AJ, Maas AI, et al. Some prognostic models for traumatic brain injury were not valid. *J Clin Epidemiol* 2006;59:132-43.
 41. Bleeker SE, Moll HA, Steyerberg EW, et al. External validation is necessary in prediction research: a clinical example. *J Clin Epidemiol* 2003;56:826-32.
 42. Altman DG, Royston P. What do we mean by validating a prognostic model? *Stat Med* 2000;19:453-73.
 43. Righini M, Bounameaux H. External validation and comparison of recently described prediction rules for suspected pulmonary embolism. *Curr Opin Pulm Med* 2004;10:345-9.
 44. Justice AC, Covinsky KE, Berlin JA. Assessing the generalizability of prognostic information. *Ann Intern Med* 1999;130:515-24.
 45. Grol R, Grimshaw J. From best evidence to best practice: effective implementation of change in patients' care. *Lancet* 2003;362:1225-30.
 46. Bero LA, Grilli R, Grimshaw JM, Harvey E, Oxman AD, Thomson MA. Closing the gap between research and practice: an overview of systematic reviews of interventions to promote the implementation of research findings. The Cochrane Effective Practice and Organization of Care Review Group. *BMJ* 1998;317:465-8.

Supplier

- a. Insightful Corp, 1700 Westlake Ave N, Ste 500, Seattle, WA 98109-3044.